MATHEMATICAL TESTS:—THEIR RELATION TO THE MATHEMATICS TEACHER.

By J. H. Minnick.

To-day we are hearing a great deal about standardized tests and scales. School surveys make extensive use of them in determining the efficiency of school systems, and bureaus of educational measurements are being established throughout the country for the purpose of applying them systematically to our schools. Such a movement can not fail to make its influence felt in the class room. The purpose of this paper is to discuss mathematical tests in their relation to the class-room teacher. The development of tests and their use for administrative and experimental purposes concern us only in so far as they affect the teacher of mathematics.

We shall first discuss some of the dangers which attend the use of tests and scales. Later we shall consider the advantages which result from their use.

The most successful mathematical scales and tests can be developed only by taking into account the contributions which mathematics should make to the realization of the general aim of education. Education does not, or at least should not, consist of independent factors each contributed by a specific subject. The contributions of the various subjects should fit into each other to form a unit. The aim of education should be definitely stated and each subject so taught that it will contribute most effectively to the realization of this aim. If then a mathematical test is to serve the purpose of education most effectively, we must be sure that it measures an ability which is essential to the aim of education. Furthermore, we must have a means of measuring every such essential ability or our subject will be in danger of making a stinted contribution to education.

Here then is one of the chief dangers connected with the use of mathematical tests. Either the makers of mathematical tests have not been conscious of the educational value of mathe-

matics, or they have been unable to develop tests for the measurement of all abilities underlying that value. Most tests deal with the mechanical or tool phases of the subject and fail to measure the more important products of mathematical education. It is true that in some cases* reasoning tests have been developed both in algebra and geometry; however, the exercises of these tests do not consist of out-of-school situations such as those to which the child will have to apply his mathematics in the future. The ability to do these tests does not imply the ability to use mathematics in the solution of real life situations. Moreover, I know of no tests that attempt to measure the child's appreciation of the relation of mathematics to his material and social environments, nor have I found an attempt to measure the child's ability to recognize a mathematical situation or his ability to use his mathematical knowledge to improve present methods of work in the shop, office, or elsewhere.

Although existing tests are thus only a partial measure of the educational products of mathematics, nevertheless superintendents and administrators are using them either consciously or unconsciously as a measure of teaching ability. If pupils of one teacher make a higher score than those of another, there is a feeling that the former is the better teacher, regardless of the unmeasured results obtained by the second teacher. If thus judged, the teacher naturally gives attention to those phrases of the subject in which her pupils are to be tested, which of course means that she will emphasize the mechanical to the neglect of the more important values of the subject. Hence, unless tests and scales are used with the greatest care there is danger of formalizing mathematics more than it ever has been and thus defeating the purpose of that better group of teachers who have been trying to vitalize the work of our secondary schools.

If, as we shall note later, a scale is to be used for the purpose of diagnosis and for determining when a given ability has been sufficiently developed, then it should, throughout all its parts,

* H. G. Hotz, "First Year Algebra Scales," pp. 8–9, 18–20. Agnes L. Rogers, "Experimental Tests of Mathematical Ability and their Prognostic Value," pp. 21, 29–31, 39–40. Indiana University Algebra Tests, Tests VII and XII. J. H. Minnick, "An Investigation of Certain Abilities Fundamental to the Study of Geometry," pp. 15–19.

measure the same ability. This ability may be dependent upon others in such a way that the latter are indirectly measured; but it is important that in passing over any two parts of the scale the ability directly involved shall vary only in quantity and not in quality. It is doubtful if this principle has been observed in the construction of certain tests. In Woody's Addition Scale* of the two exercises $\begin{matrix}23\\18\\25\end{matrix}$ and $25+42$ the latter is assigned the greater value. Is it true that the solution of the second exercise involves exactly the same kind of ability as the first but to a greater degree? Is there no new difficulty involved in the second, calling into play a new type of ability? Must a child be able to solve satisfactorily problems of the second type before he can satisfactorily solve the first type? Again in the same scale we find three exercises arranged according to their relative difficulties as follows:

(33)	(34)	(35)
.49	1/6 + 3/8	2 ft. 6 in.
.28		3 ft. 5 in.
.63		4 ft. 9 in.
.95		
1.69		
.22		
.33		
.36		
1.01		
.56		
.88		
.75		
.56		
1.10		
.18		
.56		

If a pupil does (33), (34) and (35) correctly, does it mean that in each succeeding case he is exercising the same ability only to a greater degree; or if he does (33) but can not do (34) or (35), does it mean that (33) marks the limit of his ability to

* Clifford Woody, "Measurements of Some Achievements in Arithmetic," pp. 4 and 16.

do problems of that type and that he should therefore have further drill on the addition of decimal fractions? Dr. G. H. Hotz* has developed one scale for addition and subtraction, another for multiplication and division, and a third for equations and formulæ. Do the two subjects composing each of these pairs involve the same ability? Do the solutions of $4r+3r+2r$ and $8c-(-6-3c)$ require the same type of ability? Does it require just a little more of the same kind of ability to pass from

$$\frac{p^2+4p-45}{p^2-2p-4}\cdot\frac{p^3-8}{p^2-81}\cdot\frac{1}{3pr-152}$$

to

$$\frac{X^2+27}{X^2-X-12}\div\frac{3X+9}{X+4}?$$

Will any amount of increase in the ability to solve

$$c-2(3-4c)=12$$

enable the pupil to solve the following:

The area of a triangle $=\frac{1}{2}\,bh$, in which $b=$ length of the base and $h=$ height of the triangle.

How many square feet are there in the area of a triangle whose base is 10 feet, and whose height is 8 feet?

Unless these questions can be answered in the affirmative, such tests, although having other important functions, can not safely be used for the purpose of diagnosis and for determining when a class has gained a satisfactory mastery of a given phase of the work.

We have referred to the tendency on the part of some to use tests as a means of comparing teaching ability. It should be definitely understood that generally neither absolute nor relative teaching ability can be determined by the use of the tests. In the first place, the norms for the various tests have been established by examining large numbers of children in our schools. Hence, these norms represent conditions which actually exist rather than those which should exist. We freely admit the inefficiency of our present school system and there-

* G. H. Hotz, "First Year Algebra Scales," pp. 5-7.

fore it is probable that these norms are not correct. Doubtless in many cases they are too low and in some cases they may be too high. For example, in algebra it may be that we devote so much time to long division, factoring, complicated fractional equations and the like, that either the degree of skill attained passes the point of usefulness or we are forced to neglect other topics which would have a greater usefulness than that secured by excessive drill on these subjects. If such conditions exist, a norm based upon them would be too high. Hence, since we do not know that the standards established for these tests are correct, it is impossible to use them as an absolute measure of a teacher's ability. Furthermore, the achievements of pupils are functions of many variables, chief among which are teaching ability, methods of instruction, standard school conditions, home influence, and the pupils' native ability. No one can say just what each of these factors should be in order to maintain a certain standard of achievement. Until this can be done we can not say to what factor a low score is due, and therefore, even if we knew that the established norms are correct, tests could not be used as an absolute measure of teaching ability.

In like manner the relative abilities of two teachers generally cannot be determined by the use of tests. To do this, the scores made by the pupils of the teachers in question would be compared. But since these scores are the results of many factors, a variation in them may be due to a variation of some factor other than teaching ability. Hence, it is impossible to compare the teaching abilities of two teachers by means of tests unless all other factors are constant, a condition which seldom or never exists.

It is highly desirable that tests or scales for measuring individual progress and for diagnosing the weakness of individual pupils should be developed. Such an impersonal standard for judging individuals would be of value in rating pupils and such a means of diagnosis would render individual instruction far more efficient. However, a pupil's reaction towards a test is conditioned by his physical condition, his mental attitude created by the test, his individual store of knowledge, and various external influences. The influence of these varying factors may be partially eliminated by the use of several equivalent tests

given at different times. But such equivalent tests have not been developed and until they have been, too much emphasis should not be placed on the scores of individual pupils. This difficulty can, however, be partially removed and existing tests be used with individuals to advantage if the teacher will check the results with supplementary tests of her own making.

So far we have considered the dangers that attend the use of tests. Nevertheless, when properly constructed and used, they should prove to be of great value to the class-room teacher. In the first place, we should realize that the tool side of mathematics is essential to the more valuable phases of the subject and we should welcome any means of measuring growth along these lines. As indicated above, it is for exactly this purpose that most of the existing tests have been developed. These tests may serve at least three important functions. In each of these cases they will be used to test groups of children and the large number of pupils will tend to eliminate the individual variations which are so troublesome when we are concerned with individual scores.

First, they may be used as an impersonal standard by which a teacher may measure the group products of her efforts. Although the established norms are based on conditions as they now exist rather than as they should be, nevertheless since they combine the judgment of many teachers, they are doubtless a safer standard than the opinion of any one teacher. Also the teacher's individual judgment is colored by her personal experience with her class, her likes and dislikes, and her individual ideals. Tests furnish an external means of judging, from which these personal elements have been eliminated.

Second, efficient teaching demands some means of determining when a class has attained a satisfactory standard. Unless this is possible, the teacher may pass from one phase of the work to another too soon, with a loss of time due to the fact that the pupils are not sufficiently prepared for the advanced work, or on the other hand, he may waste quite as much time by dwelling on a topic after the necessary standard has been reached. Standardized tests, properly developed, would thus enable the teacher to better economize his time. This implies that he knows exactly what ability each test measures and also

that he knows and does not neglect those other abilities for which tests have not been developed.

Finally, tests and scales may be used for the purpose of diagnosis. Teachers frequently find that their class is doing poor work, but are unable to determine the exact difficulty. If a complete analysis were made of the abilities involved in the study of a given subject such as geometry and a test were developed for the measurement of each of these abilities, then by an application of these tests the teacher could determine the exact cause of the failure and remove the difficulty with a greater degree of efficiency. Obviously, no such analysis has been made and satisfactory tests have not been developed for all known essential abilities; but a start has been made and we should not hesitate to use any test which promises aid in the diagnosis of class weaknesses. However, we must remember that tests have not been developed for all abilities necessary to the study of mathematics, and that difficulties may exist which none of the tests thus far developed will reveal.

In conclusion, it is evident that scale building is far from complete. More work has been done in mathematics than in any other secondary field. Yet, there is much to be done before we have a complete and satisfactory set of scales for our own subject. Those scales which have been developed have many imperfections and for the most part they deal with the formal aspect of mathematics. However, they constitute a step in the right direction, and although they are only a partial measure of the contribution which mathematics should make to education, nevertheless, when properly used their value to the class-room teacher can scarcely be overestimated.

UNIVERSITY OF PENNSYLVANIA,
PHILADELPHIA, PA.